

Chapter 19

Fields and geometry

Introduction

There is a story mathematicians often tell about the construction of new numbers from old which goes like this. To construct the negative integers from the positive integers, we solve equations like $x + 1 = 0$; to construct the rational numbers from the integers we solve equations like $mx = n$, where m and n are integers; and to construct the complex numbers from the real numbers we solve equations like $x^2 = -1$. (The real numbers do not fit into this story.) This is not an accurate account of how these types of numbers were discovered, but it makes a substantial mathematical point: faced with a problem you have to solve, it can be useful to create objects that count as the answers. In all of these cases the ‘new’ numbers can be handled by analogy with the old ones.

In this chapter, we look at two aspects of this process. Firstly, we consider the numbers that arise as the roots of polynomial equations of any degree and with rational coefficients, that is, equations of the form

$$a_0x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n = 0,$$

where $a_i \in \mathbb{Q}$ for $i = 0, \dots, n$. These numbers are called *algebraic numbers*, and we show that they form a field. Moreover, we give a standard technique for finding a polynomial equation satisfied by an algebraic number. The basic building blocks of the corresponding theory will be developed in the first half of this chapter.

Secondly, we turn to the algebraic version of a story that has its roots in the classical geometry of ancient Greece. Around the time of Euclid (the second and third century BC), geometric problems often involved constructions. A famous one asked for the construction of a square twice the area of a given square, and another asked for a method to divide a given line segment into a given number of equal subsegments.

Both these problems can be solved by the use of a straightedge (an unmarked ruler) and a compasses (for drawing circles). Difficulties, however, arose with analogous problems, such as the construction of a cube twice the volume of a given cube, and the problem of dividing a given circular arc (or, equivalently, a given angle) into three equal parts. Although the Greeks solved these problems geometrically, their solutions involved elaborate methods such as the use of conic sections, and no solutions were found using straightedge and compasses alone. When these two problems were revived in the modern West around the time of Descartes in the seventeenth century, it came to be believed that no solutions could be found by straightedge and compasses alone, and this was eventually shown to be the case by an otherwise little-known French mathematician, Pierre Wantzel, in 1837.

Alone in human investigation, mathematics offers the chance to prove that something is impossible; the other sciences can merely establish that something is very unlikely. But what allows mathematicians to prove that the cube cannot be doubled and the angle cannot be trisected by the required methods?

One clue comes from expressing geometric problems in algebraic terms. To find the side of cube that is double the volume of another (say, a cube of unit volume) is to find an x such that $x^3 = 2$. To express the problem of trisecting an angle of φ in algebraic terms we first recall the trigonometrical formula

$$\cos 3\theta = 4 \cos^3 \theta - 3 \cos \theta.$$

If we set $3\theta = \varphi$ this equation becomes

$$\cos \varphi = 4 \cos^3 \left(\frac{\varphi}{3} \right) - 3 \cos \left(\frac{\varphi}{3} \right),$$

which we regard as a cubic equation for $x = \cos(\frac{\varphi}{3})$ given the value, a say, of $\cos \varphi$:

$$4x^3 - 3x - a = 0.$$

On the other hand, when we make a geometric construction in the plane we know a line segment if, and only if, we know the coordinates of its endpoints. This means that we know all the line segments whose endpoints have rational coordinates: start from the segment that goes from the origin to the point $(1,0)$ and copy it (by shifting and rotating) and divide it (using similar triangles) as necessary.

Thus, we can draw lines with rational slopes that pass through a point with rational coordinates by joining pairs of known points. Likewise, we can draw circles with their centres at points with rational coordinates and that have radii whose squares are rational, by taking one such point as the centre and the radius as a line segment from the first point to some other known point.

When one of these circles meets one of these lines we can get new points, whose coordinates simultaneously satisfy the equation of the line and the equation of the circle. As you might expect, the coordinates of these points generally involve square roots, because the equation of the circle contains squared terms and so the equations to be solved may well be quadratic equations. A similar result occurs when one of these circles meets another of these circles.

Now, a geometric construction involves a sequence of operations each of which involves drawing either a circle or a line, and at each stage the only points available are the ones already constructed. So at each stage we construct points whose coordinates are either rational numbers, or square roots of rational numbers, or square roots of numbers that involve square roots of rational numbers, and so on. At each stage the new coordinates are obtained by solving a family of linear and quadratic equations, and the roots might be rational, or roots of an equation of degree 2, or roots of an equation of degree 4, and so on.

But a cubic equation is surely unlikely to arise in this way, and if we can show that no (irreducible) cubic equation arises, then we have indeed proved that the cube cannot be doubled in volume, nor the angle trisected, by straightedge and compasses alone, all because 3 is not a power of 2.

The second half of this chapter is given over to establishing how such proofs can be achieved.

Chapter guide

The first section of this chapter revises and develops the material of Chapter 18. It will require reasonable fluency at working with polynomials. You will also need to be able to carry out Gaussian elimination to find the determinant of a matrix. There is a worked exercise that should help you with this.

The second section is about ruler and compasses constructions and involves elementary geometry. You may find it helpful to actually do the constructions using paper and pencil. The text states key results from geometry as they are needed.

1 Field extensions revisited

The field extensions introduced and discussed in Chapter 18 were obtained by adjoining a finite number of elements to the base field – usually, just a single element. The solution to the first exercise of this chapter will enable us to generalise this construction to extensions by arbitrary sets of elements.

Exercise 1.1

Let F be a subfield of a field E and let S be a subset of elements of E . Let L be the intersection $\bigcap M_i$ of all the subfields M_i of E that contain both S and F . Show that

- (a) L is a subfield of E
- (b) if K is a subfield of E with $F \subseteq K$ and $S \subseteq K$ then L is a subfield of K .

Note that the index i in M_i is for notational convenience. You should not assume that it is an integer as there could be an uncountably infinite number of such fields.

Exercise 1.1 allows us to define the subfield generated by a set.

Definition 1.1 Subfield generated by a set

Let F be a subfield of a field E and let S be a subset of elements of E . We define $F(S)$ to be the smallest subfield of E containing F and all the elements of S . If no field F is specified, then the smallest subfield of E that contains S is called the subfield of E **generated** by S .

Exercise 1.1 tells us that $F(S) = \bigcap M_i$, where the M_i range over all the subfields of E that contain both S and F .

In order to match the notation introduced in Chapter 18, when $S = \{s_1, \dots, s_k\}$ is a finite set, we write $F(S) = F(s_1, \dots, s_k)$ instead of $F(\{s_1, \dots, s_k\})$.

Our aim is to show that finite-dimensional extensions are necessarily algebraic. We start by proving a simple necessary and sufficient condition for an element to be algebraic.

Theorem 1.2

Let E be an extension of a field F . An element α of E is algebraic over F if, and only if, the degree $[F(\alpha) : F]$ is finite.

Proof Let the degree $[F(\alpha) : F]$ be finite. This implies that the vector space $F(\alpha)$ has a finite dimension n , say, over F . Consequently, any set of $n + 1$ elements of K must be linearly dependent over F . Consider a particular set of consecutive powers of α for such a set, namely, $\{1, \alpha, \alpha^2, \dots, \alpha^n\}$. Since this set of $n + 1$ elements is linearly dependent, there exist coefficients $c_0, c_1, c_2, \dots, c_n \in F$, at least one of which is non-zero, such that

$$c_0 + c_1\alpha + c_2\alpha^2 + \dots + c_n\alpha^n = 0.$$

Since at least one coefficient c_j is non-zero, the above equation shows that α is a root of a non-zero polynomial over F .

Conversely, let $\alpha \in E$ be a root of a polynomial over F . Then, by Theorem 2.21(c) of Chapter 18, α is a root of an irreducible polynomial over F and the degree $[F(\alpha) : F]$ is finite (and equal to the degree of the irreducible polynomial that has α as a root). ■

Exercise 1.2

Show that if F is a subfield of E and $\alpha \in E$ is algebraic over F and if K is any field such that $F \subset K \subset E$, then α is also algebraic over K .

We are now ready to prove a main result that characterises field extensions of finite degree.

Theorem 1.3 *Field extensions of finite degree*

Let $F \subset E$ be fields. Then the degree $[E : F]$ is finite if, and only if, E is an algebraic extension of F and there exists a finite subset S of E such that $E = F(S)$.

Proof Suppose that $[E : F]$ is finite, say equal to n . This implies that E , regarded as a vector space over the field F , has a basis $S = \{\beta_1, \beta_2, \dots, \beta_n\}$ consisting of elements of E . Now $F \subseteq E$ and $S \subseteq E$ and so $F(S) \subseteq E$. Since every element of E can be expressed in the form $c_1\beta_1 + c_2\beta_2 + \dots + c_n\beta_n$ for some $c_1, c_2, \dots, c_n \in F$, it is also the case that $E \subseteq F(S)$ because $F(S)$ is a field and so closed under addition. Hence $E = F(S)$.

Suppose that α is an arbitrary element of E . Then $F(\alpha) \subset E$ and so, by Theorem 2.29 of Chapter 18, $[F(\alpha) : F] \leq [E : F] = n$. This implies that the degree $[F(\alpha) : F]$ is finite and so α is algebraic over F by Theorem 1.2.

Conversely, suppose that E is an algebraic extension of F and that there is $S = \{s_1, s_2, \dots, s_k\}$ with $E = F(S)$. Since E is an algebraic extension, all the elements in S are algebraic over F and hence, by the solution to Exercise 1.2, they are algebraic over any field L such that $F \subset L \subset E$.

From Theorem 1.2 we know that the degrees $[L(s_j) : L]$ are finite for any j such that $1 \leq j \leq k$, and for any field L such that $F \subset L \subset E$. It follows that if $L_j = F(s_1, \dots, s_{j-1})$ for $1 \leq j \leq k$, then the degrees $[L_j(s_j) : L_j]$ are all finite.

We now apply Theorem 2.29 of Chapter 18 to the chain of fields

$$F \subseteq F(s_1) \subseteq F(s_1, s_2) \subseteq \dots \subseteq F(S) = E$$

to obtain

$$[E : F] = [F(s_1) : F][F(s_1, s_2) : F(s_1)] \cdots [F(S) : F(s_1, s_2, \dots, s_{k-1})].$$

Since all the factors on the right-hand side are finite, we conclude that $[E : F]$ is finite. ■

1.1 Algebraic extensions revisited

The next two subsections develop some techniques for doing calculations in algebraic extensions. This subsection starts with an exercise that extends the idea of ‘rationalisation’ to help find inverses of elements in $\mathbb{Q}(\sqrt{2}, \sqrt{3})$. Consider an element $a + b\sqrt{3} \in \mathbb{Q}(\sqrt{2}, \sqrt{3})$ with $a, b \in \mathbb{Q}(\sqrt{2})$. We can write

$$\frac{1}{a + b\sqrt{3}} = \frac{a - b\sqrt{3}}{a^2 - 3b^2}$$

by multiplying both the numerator and the denominator of the fraction on the left-hand side by $a - b\sqrt{3}$. The important observation is that the denominator of the fraction on the right-hand side now only involves elements of $\mathbb{Q}[\sqrt{2}]$. We can repeat the process by multiplying top and bottom by appropriate expressions so that the denominator eventually becomes an element of $\mathbb{Q}$.

Exercise 1.3

Let $K = \mathbb{Q}(\sqrt{2}, \sqrt{3})$.

- (a) By writing $1 + \sqrt{2} + 2\sqrt{3} + \sqrt{6}$ as $1 + \sqrt{2} + (2 + \sqrt{2})\sqrt{3}$, show that

$$\frac{1}{1 + \sqrt{2} + 2\sqrt{3} + \sqrt{6}} = \frac{1 + \sqrt{2} - 2\sqrt{3} - \sqrt{6}}{-15 - 10\sqrt{2}}.$$

- (b) Find the inverse of $1 + \sqrt{2} + 2\sqrt{3} + \sqrt{6}$ in the form $a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6}$ where $a, b, c, d \in \mathbb{Q}$.

The next worked exercise and the following exercise explore the field $\mathbb{Q}(\sqrt{2}, \sqrt{3})$.

Worked Exercise 1.4 *A polynomial over $\mathbb{Q}$ with root $\sqrt{2} + \sqrt{3}$*

Consider the field $K = \mathbb{Q}(\sqrt{2}, \sqrt{3})$. In the solution to Exercise 2.10 of Chapter 18 we saw that $[K : \mathbb{Q}] = 4$. By Theorem 1.3 every element of K is algebraic over $\mathbb{Q}$. This implies that the element $\sqrt{2} + \sqrt{3}$ is itself algebraic over $\mathbb{Q}$. Since the degree $[K : \mathbb{Q}]$ is 4, the number $\sqrt{2} + \sqrt{3}$ is a root of some polynomial of degree at most 4 over $\mathbb{Q}$. Find such a polynomial.

Solution

We first note that $(\sqrt{2} + \sqrt{3})^2 = 5 + 2\sqrt{6}$ and that $(5 + 2\sqrt{6})(5 - 2\sqrt{6}) = 1 \in \mathbb{Q}$. The latter observation gives

$$(y - (5 + 2\sqrt{6}))(y - (5 - 2\sqrt{6})) = y^2 - 10y + 1.$$

Thus a suitable polynomial is $x^4 - 10x^2 + 1$.

Since a factorisation in $\mathbb{R}$ is $(x^2 - (5 + 2\sqrt{6}))(x^2 - (5 - 2\sqrt{6}))$ and the factors are not in $\mathbb{Q}[x]$, this polynomial is irreducible over $\mathbb{Q}$. Since it is monic, by Exercise 2.2 in Chapter 18, this is the minimal polynomial for $\sqrt{2} + \sqrt{3}$ over $\mathbb{Q}$.

In Worked Exercise 1.4 we found a polynomial equation over the field $\mathbb{Q}$ for $\sqrt{2} + \sqrt{3}$ by using specific algebraic properties of this element. In the next subsection we will develop a reasonably fast method for determining a polynomial over a given field with a root outside the field that does not depend on such algebraic properties.

Exercise 1.4

- (a) Determine all the roots of the polynomial $x^4 - 10x^2 + 1$.
 (b) Show that $\mathbb{Q}(\sqrt{2}, \sqrt{3}) = \mathbb{Q}(\sqrt{2} + \sqrt{3})$.
 (c) Show that this field is the splitting field of the irreducible polynomial $x^4 - 10x^2 + 1$.

1.2 The field of algebraic numbers

Suppose that a field E is an extension of a field F . Our next aim is to study the collection of all elements of E that are algebraic over F . We start by proving that this set is closed under addition and multiplication.

Theorem 1.5

Let $F \subset E$ be fields and let $\alpha, \beta \in E$ be algebraic over F . Then both $\alpha + \beta$ and $\alpha\beta$ are elements of E that are algebraic over F .

Proof If α and β are algebraic over F , then β is also algebraic over $F(\alpha)$ by Exercise 1.2. Applying Theorem 2.29 of Chapter 18 to the chain of fields

$$F \subset F(\alpha) \subset F(\alpha, \beta)$$

gives

$$[F(\alpha, \beta) : F] = [F(\alpha, \beta) : F(\alpha)][F(\alpha) : F].$$

Since both factors on the right-hand side are finite, the degree $[F(\alpha, \beta) : F]$ is finite. By Theorem 1.3, every element in $F(\alpha, \beta)$ must be algebraic over F . But $\alpha + \beta$ and $\alpha\beta$ both belong to $F(\alpha, \beta)$, and hence they are algebraic. ■

One consequence of Theorem 1.5 is that whenever α and β are roots of a polynomial over a field F , the elements $\alpha + \beta$ and $\alpha\beta$ are also roots of some polynomial over F . In Example 1.4 you saw a way of finding such a polynomial where $F = \mathbb{Q}$, $\alpha = \sqrt{2}$ and $\beta = \sqrt{3}$. We now address the question of doing this in full generality.

The idea is as follows. Let $F \subseteq E$ with $[E : F] = n$ and let $\alpha \in E \setminus F$. Then the $n + 1$ elements $1, \alpha, \alpha^2, \dots, \alpha^n$ are linearly dependent in the n -dimensional vector space E over the field F . The linear dependence means that there are $n + 1$ elements $c_0, c_1, c_2, \dots, c_n \in F$, not all zero, such that

$$c_0 1 + c_1 \alpha + c_2 \alpha^2 + \dots + c_n \alpha^n = 0,$$

from which a polynomial equation for α can be extracted.

However, it may not be obvious to you at this moment how this idea works in practice for more complicated algebraic elements, such as, for instance, $\alpha = \sqrt{2} + \sqrt[3]{2}$. The next worked example illustrates this approach for this particular element. The method avoids direct calculations of high powers of α and, at the same time, it provides an interesting application of certain elementary facts from linear algebra regarding linear dependence, homogeneous linear systems and determinants (known to you from earlier modules, albeit only over the field of real numbers).

Worked Exercise 1.6

Find a polynomial over $\mathbb{Q}$ with root $\sqrt{2} + \sqrt[3]{2}$.

Solution

By Theorem 1.5, the number $\alpha = \sqrt{2} + \sqrt[3]{2}$ is algebraic over $\mathbb{Q}$. Now $[\mathbb{Q}(\sqrt{2}) : \mathbb{Q}] = 2$ and $\sqrt[3]{2}$ is a root of $x^3 - 2$. Since $x^3 - 2$ is monic and irreducible, it is the minimal polynomial for $\sqrt[3]{2}$ over $\mathbb{Q}$, and so any finite field extension of $\mathbb{Q}$ containing this root must have degree a multiple of 3. Thus $[\mathbb{Q}(\sqrt{2}, \sqrt[3]{2}) : \mathbb{Q}(\sqrt{2})] = 3$ and so by Theorem 2.29 from Chapter 18 (the KLM Theorem), we have $[\mathbb{Q}(\sqrt{2}, \sqrt[3]{2}) : \mathbb{Q}]$ is 6.

Hence, since $\alpha \in \mathbb{Q}(\sqrt{2}, \sqrt[3]{2})$, the element α is a root of a polynomial over $\mathbb{Q}$ of degree at most 6. As in the proof of the KLM Theorem, a basis for the vector space $\mathbb{Q}(\sqrt{2}, \sqrt[3]{2})$ over the field $\mathbb{Q}$ is $B = \{u_1, u_2, \dots, u_6\}$, where

$$u_1 = 1, u_2 = \sqrt[3]{2}, u_3 = \sqrt[3]{4}, u_4 = \sqrt{2}, u_5 = \sqrt{2} \cdot \sqrt[3]{2}, u_6 = \sqrt{2} \cdot \sqrt[3]{4}.$$

Now, every element of E can be expressed as a linear combination of elements of B with coefficients in F in a unique way. In particular, this applies to the six elements $u_1\alpha, u_2\alpha, \dots, u_6\alpha$. Our first step is to find these linear combinations, which we do by direct calculation.

$$\begin{aligned} u_1\alpha &= \sqrt{2} + \sqrt[3]{2} \\ u_2\alpha &= \sqrt{2}\sqrt[3]{2} + \sqrt[3]{4} \\ u_3\alpha &= \sqrt{2}\sqrt[3]{4} + 2 \\ u_4\alpha &= 2 + \sqrt{2}\sqrt[3]{2} \\ u_5\alpha &= 2\sqrt[3]{2} + \sqrt{2}\sqrt[3]{4} \\ u_6\alpha &= 2\sqrt[3]{4} + 2\sqrt{2} \end{aligned}$$

Writing the right-hand sides of these equations in terms of $u_1, \dots, u_6$ yields the following linear system.

$$\begin{aligned} u_1\alpha &= 0 \cdot u_1 + 1 \cdot u_2 + 0 \cdot u_3 + 1 \cdot u_4 + 0 \cdot u_5 + 0 \cdot u_6 \\ u_2\alpha &= 0 \cdot u_1 + 0 \cdot u_2 + 1 \cdot u_3 + 0 \cdot u_4 + 1 \cdot u_5 + 0 \cdot u_6 \\ u_3\alpha &= 2 \cdot u_1 + 0 \cdot u_2 + 0 \cdot u_3 + 0 \cdot u_4 + 0 \cdot u_5 + 1 \cdot u_6 \\ u_4\alpha &= 2 \cdot u_1 + 0 \cdot u_2 + 0 \cdot u_3 + 0 \cdot u_4 + 1 \cdot u_5 + 0 \cdot u_6 \\ u_5\alpha &= 0 \cdot u_1 + 2 \cdot u_2 + 0 \cdot u_3 + 0 \cdot u_4 + 0 \cdot u_5 + 1 \cdot u_6 \\ u_6\alpha &= 0 \cdot u_1 + 0 \cdot u_2 + 2 \cdot u_3 + 2 \cdot u_4 + 0 \cdot u_5 + 0 \cdot u_6 \end{aligned}$$

Subtracting $u_i\alpha$ from both sides of each equation gives a homogeneous linear system of equations with the 6×6 coefficient matrix

$$\begin{pmatrix} -\alpha & 1 & 0 & 1 & 0 & 0 \\ 0 & -\alpha & 1 & 0 & 1 & 0 \\ 2 & 0 & -\alpha & 0 & 0 & 1 \\ 2 & 0 & 0 & -\alpha & 1 & 0 \\ 0 & 2 & 0 & 0 & -\alpha & 1 \\ 0 & 0 & 2 & 2 & 0 & -\alpha \end{pmatrix}.$$

Since we know that α is in the field and that α is not zero, we have a non-zero solution of this system, namely, the 6-tuple $u_1, \dots, u_6$. This implies that the determinant of the matrix of this system must be equal to zero that is,

$$\begin{vmatrix} -\alpha & 1 & 0 & 1 & 0 & 0 \\ 0 & -\alpha & 1 & 0 & 1 & 0 \\ 2 & 0 & -\alpha & 0 & 0 & 1 \\ 2 & 0 & 0 & -\alpha & 1 & 0 \\ 0 & 2 & 0 & 0 & -\alpha & 1 \\ 0 & 0 & 2 & 2 & 0 & -\alpha \end{vmatrix} = 0.$$

Evaluating the determinant in terms of α gives a polynomial expression in terms of α , which in turn gives a polynomial equation satisfied by α .

We now perform elementary row and column operations to evaluate the determinant. Although some row operations do change the value of the determinant, by either changing the sign or multiplying by a constant, a zero determinant is not affected by any of these operations. We first observe that the top right-hand block of the determinant has the form of a 3×3 identity matrix. As a result, our first step swaps the three right-hand columns with the three left-hand ones. We then proceed with Gaussian elimination.

First we swap the first three columns:

$$\begin{vmatrix} 1 & 0 & 0 & -\alpha & 1 & 0 \\ 0 & 1 & 0 & 0 & -\alpha & 1 \\ 0 & 0 & 1 & 2 & 0 & -\alpha \\ -\alpha & 1 & 0 & 2 & 0 & 0 \\ 0 & -\alpha & 1 & 0 & 2 & 0 \\ 2 & 0 & -\alpha & 0 & 0 & 2 \end{vmatrix} = 0.$$

Next we clear the first column below the diagonal:

$$\left| \begin{array}{cccccc} 1 & 0 & 0 & -\alpha & 1 & 0 \\ 0 & 1 & 0 & 0 & -\alpha & 1 \\ 0 & 0 & 1 & 2 & 0 & -\alpha \\ 0 & 1 & 0 & 2 - \alpha^2 & \alpha & 0 \\ 0 & -\alpha & 1 & 0 & 2 & 0 \\ 0 & 0 & -\alpha & 2\alpha & -2 & 2 \end{array} \right| \quad \begin{array}{l} r_4 = r_4 + \alpha r_1 \\ r_6 = r_6 - 2r_1. \end{array}$$

Now we clear the second column below the diagonal:

$$\left| \begin{array}{cccccc} 1 & 0 & 0 & -\alpha & 1 & 0 \\ 0 & 1 & 0 & 0 & -\alpha & 1 \\ 0 & 0 & 1 & 2 & 0 & -\alpha \\ 0 & 0 & 0 & 2 - \alpha^2 & 2\alpha & -1 \\ 0 & 0 & 1 & 0 & 2 - \alpha^2 & \alpha \\ 0 & 0 & -\alpha & 2\alpha & -2 & 2 \end{array} \right| \quad \begin{array}{l} r_4 = r_4 - r_2 \\ r_5 = r_5 + \alpha r_2. \end{array}$$

Clearing the third column below the diagonal gives

$$\left| \begin{array}{cccccc} 1 & 0 & 0 & -\alpha & 1 & 0 \\ 0 & 1 & 0 & 0 & -\alpha & 1 \\ 0 & 0 & 1 & 2 & 0 & -\alpha \\ 0 & 0 & 0 & 2 - \alpha^2 & 2\alpha & -1 \\ 0 & 0 & 0 & -2 & 2 - \alpha^2 & 2\alpha \\ 0 & 0 & 0 & 4\alpha & -2 & 2 - \alpha^2 \end{array} \right| \quad \begin{array}{l} r_5 = r_5 - r_3 \\ r_6 = r_6 + \alpha r_3. \end{array}$$

At this stage we stop as we can see that the equation we are seeking is equivalent to

$$\left| \begin{array}{ccc} 2 - \alpha^2 & 2\alpha & -1 \\ -2 & 2 - \alpha^2 & 2\alpha \\ 4\alpha & -2 & 2 - \alpha^2 \end{array} \right| = 0.$$

Evaluating this 3×3 determinant gives

$$\begin{aligned} & (2 - \alpha^2)[(2 - \alpha^2)^2 + 4\alpha] - 2\alpha[-2(2 - \alpha^2) - 8\alpha^2] \\ & - [4 - 4\alpha(2 - \alpha^2)] = 0, \end{aligned}$$

which simplifies to

$$\alpha^6 - 6\alpha^4 - 4\alpha^3 + 12\alpha^2 - 24\alpha - 4 = 0.$$

It follows that $\alpha = \sqrt{2} + \sqrt[3]{2}$ is a root of the polynomial

$$x^6 - 6x^4 - 4x^3 + 12x^2 - 24x - 4.$$

We could have carried out further row operations to reduce our calculation to the evaluation of a 2×2 determinant, but it seemed easier to stop and evaluate the determinant directly.

The following strategy summarises the approach taken in Worked Exercise 1.6.

Strategy: to find a polynomial over F that has α as a root

Let E be an extension of a field F such that $[E : F] = n$ and let α be an element of E .

1. Write down a basis $B = \{u_1, \dots, u_n\}$ for E as a vector space over F .
2. For $j = 1, \dots, n$, express $u_j\alpha$ as a linear combination of elements of B , that is

$$u_j\alpha = c_{j,1}u_1 + \dots + c_{j,j}u_j + \dots + c_{j,n}u_n.$$

3. Rearrange the equations obtained in step 2 as

$$c_{j,1}u_1 + \dots + (c_{j,j} - \alpha)u_j + \dots + c_{j,n}u_n = 0.$$

4. Write down the coefficient matrix A of the resulting system of n homogeneous linear equations.
5. Since $u_1, u_2, \dots, u_n$ is a non-zero solution for this system of equations, we know that $\det A = 0$.
6. The determinant $\det A$ will be a polynomial of degree at most n in α . This is the required polynomial.

Exercise 1.5

- (a) Find a polynomial equation satisfied by $1 + \sqrt{2}$ in the field $\mathbb{Q}(\sqrt{2})$.
- (b) Find a polynomial equation satisfied by $\sqrt{2} - \sqrt{3}$ in $\mathbb{Q}(\sqrt{2}, \sqrt{3})$.

Theorem 1.5 shows that the sum and the product of two algebraic elements over a field is again algebraic over the same field. In fact, the set of algebraic elements forms a subfield, as we will now prove.

Theorem 1.7

Let E and F be two fields such that $F \subset E$, and let A be the set of all the elements of E that are algebraic over F . Then A is a subfield of E .

Proof We will use the subfield criterion (Book C, Chapter 11, Lemma 1.21). By Theorem 1.5, for any $\alpha, \beta \in A$ we have $\alpha + \beta \in A$ and $\alpha\beta \in A$. In particular, from the latter we also see that $-\beta = (-1)\beta \in A$, and so $\alpha - \beta \in A$.

Let $\alpha \neq 0$ be a root of a polynomial $a_0 + a_1x + \cdots + a_{n-1}x^{n-1} + a_nx^n$ where $a_0, \dots, a_n \in F$, so that

$$a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1} + a_n\alpha^n = 0.$$

Then, since $\alpha^{-1} \in E$, taking out $\alpha^n \neq 0$ gives

$$\alpha^n(a_n + a_{n-1}\alpha^{-1} + \cdots + a_1(\alpha^{-1})^{n-1} + a_0(\alpha^{-1})^n) = 0.$$

Thus α^{-1} is a root of the polynomial $a_n + a_{n-1}x + \cdots + a_1x^{n-1} + a_0x^n$ and so $\alpha^{-1} \in A$.

Since the subset A of E is closed under the operations of addition and multiplication in E , contains 0 and 1, and is closed under taking additive and multiplicative inverses, the subfield criterion tells us that A is a subfield of the field E . ■

By the Fundamental Theorem of Algebra, any polynomial over $\mathbb{Q}$ has all its roots within $\mathbb{C}$, the field of complex numbers. Therefore, any algebraic number over $\mathbb{Q}$ is a complex number, but in fact we do not need to refer to $\mathbb{C}$ when we describe algebraic numbers over $\mathbb{Q}$; this observation and Theorem 1.7 enable us to make the following definition.

Definition 1.8 Field of algebraic numbers

Let E be an extension of a field F . The subfield of E that consists of all the elements of E that are algebraic over F is called the **field of algebraic numbers in E over F** and it is denoted by $A_E(F)$.

When $E = \mathbb{C}$ and $F = \mathbb{Q}$, this subfield is denoted by $\overline{\mathbb{Q}}$ and it is called the **field of algebraic numbers over $\mathbb{Q}$** .

The field $\overline{\mathbb{Q}}$ of algebraic numbers over $\mathbb{Q}$ is one of the most important fields that appear in mathematics. By methods well beyond the scope of this module, it can be shown that the field $\overline{\mathbb{Q}}$ has countably many elements and that the degree $[\overline{\mathbb{Q}} : \mathbb{Q}]$ is infinite. Other results about this field play an important role in the relatively recent proof of Fermat's Last Theorem (Book C, Chapter 12, Section 1), which states that there are no non-trivial integral solutions to the equation $x^n + y^n = z^n$ for any $n \geq 3$.

1.3 Transcendental extensions revisited

Theorem 2.31 of Chapter 18 proves that every field F admits an extension containing a transcendental element over F , that is, an element which is not a root of any polynomial with coefficients in F . Recall that if α is a transcendental element over F , then the field $F(\alpha)$ is called a *transcendental extension* of F . We defined transcendental extensions in Definition 2.16 of Chapter 18. We restate the definition here in the following form.

Definition 1.9 *Transcendental extension*

An extension E of a field F is said to be **transcendental** if E contains at least one element that is transcendental over F .

Now if E is a transcendental extension of F , then the degree $[E : F]$ must be infinite, by Theorem 1.3. This necessary condition, however, is not sufficient. For instance, the field $\overline{\mathbb{Q}}$ of complex algebraic numbers has infinite degree over $\mathbb{Q}$, yet $\overline{\mathbb{Q}}$ is an algebraic extension of the field $\mathbb{Q}$.

A proof that a particular element $\alpha \in E$ is transcendental over F , that is, that there is no polynomial over F that has α as one of its roots, is usually hard. On the other hand, the existence of transcendental elements can, in many cases, be established easily by comparing cardinalities. For example, as we have seen, the subfield $\overline{\mathbb{Q}}$ of $\mathbb{C}$ consisting of all algebraic numbers over $\mathbb{Q}$ is countable, while the field $\mathbb{C}$ is uncountable. It follows that there are uncountably many complex numbers that are transcendental over $\mathbb{Q}$.

Theorem 2.30 of Chapter 18 shows that if α is transcendental over F , then $F(\alpha)$ is isomorphic to the field of fractions of $F[x]$. We are going to use this result to construct more examples of transcendental numbers over $\mathbb{Q}$, and, more generally, over any given field.

Theorem 1.10

Let $L(x)$ be the field of fractions of a polynomial ring $L[x]$ over a field L . Suppose that $z \in L(x)$ and $z \notin L$. Then z is transcendental over L .

Proof We know that the field L is embedded into the polynomial ring $L[x]$. From the construction of the field of fractions of an integral domain given in Section 1 of Chapter 17, we also know that $L[x]$ is embedded in the field of fractions $L(x)$. As a result, the field L is embedded into $L(x)$ by identifying any element $c \in L$ with the fraction $\frac{c}{1}$. It follows that an element $\frac{f(x)}{g(x)} \in L(x)$, where $g(x) \neq 0$, belongs to L if, and only if, $f(x) = cg(x)$ for some $c \in L$.

Let $z = \frac{f(x)}{g(x)} \in L(x)$, with $g(x) \neq 0$, and suppose that $z \notin L$. Then $f(x) \neq cg(x)$ for any $c \in L$. (Note that this condition is automatically satisfied if the polynomials $f(x)$ and $g(x)$ do not have the same degree.) We suppose that z is algebraic over L and derive a contradiction from this assumption.

By Theorem 4.2 of Chapter 18, there exists an extension K of L containing all the roots of both $f(x)$ and $g(x)$.

By the solution to Exercise 1.4 of Chapter 17, we can choose polynomials $f_1(x)$ and $g_1(x)$ in $K[x]$ so that $z = \frac{f_1(x)}{g_1(x)}$, and so that the sum of the degrees of $f_1(x)$ and $g_1(x)$ is as small as possible, that is, so that $f_1(x)$ and $g_1(x)$ do not have a common root in K .

Recall that we are assuming that z is algebraic over L . Then, by the solution to Exercise 1.2, the element z is also algebraic over K . By Theorem 2.21 of Chapter 18, this implies that there exists a non-zero irreducible polynomial $h(x)$ in the ring $K[x]$ such that $h(z) = 0$.

Let $h(x) = a_0 + a_1x + \cdots + a_{m-1}x^{m-1} + a_mx^m$, where $a_j \in K$ for all $j \in \{0, 1, \dots, m\}$ and where $a_m \neq 0$ and $m \geq 1$. Since $h(x)$ is irreducible, we also have that $a_0 \neq 0$.

Substituting $z = \frac{f_1(x)}{g_1(x)}$ in the equation $h(z) = 0$ yields

$$a_0 + a_1 \frac{f_1(x)}{g_1(x)} + \cdots + a_{m-1} \left(\frac{f_1(x)}{g_1(x)} \right)^{m-1} + a_m \left(\frac{f_1(x)}{g_1(x)} \right)^m = 0.$$

Multiplying both sides of this equation by $(g_1(x))^m$ removes the denominators and gives the equation

$$a_0(g_1(x))^m + a_1f_1(x)(g_1(x))^{m-1} + \cdots + a_{m-1}(f_1(x))^{m-1}g_1(x) + a_m(f_1(x))^m = 0$$

in the ring $K[x]$.

By rearranging this equation as

$$a_0(g_1(x))^m = -(a_1f_1(x)(g_1(x))^{m-1} + \cdots + a_{m-1}(f_1(x))^{m-1}g_1(x) + a_m(f_1(x))^m),$$

we can see that $f_1(x)$ is a divisor of $(g_1(x))^m$. Similarly, rearranging the equation as

$$a_m(f_1(x))^m = -(a_0(g_1(x))^m + a_1f_1(x)(g_1(x))^{m-1} + \cdots + a_{m-1}(f_1(x))^{m-1}g_1(x)),$$

shows that $g_1(x)$ is a divisor of $(f_1(x))^m$.

Since both $f_1(x)$ and $g_1(x)$ have all their roots in K , this implies that $f_1(x)$ and $g_1(x)$ must have the same set of roots. But we chose $f_1(x)$ and $g_1(x)$ not to have common roots. These conclusions can only hold simultaneously if $f_1(x)$ and $g_1(x)$ are both polynomials of degree zero over K , and hence also over L . In particular, the element $z = \frac{f_1(x)}{g_1(x)}$ must then be equal to some $c \in L$. This contradicts our initial assumption on z .

The contradiction arose from supposing that z is algebraic over L , and so we can conclude that z is transcendental over L . ■

One consequence of this result is that in the transcendental extension $\mathbb{Q}(\pi)$, all fractions of the form $\frac{f(\pi)}{g(\pi)}$, where $f(x)$ and $g(x)$ are polynomials with rational coefficients and $g(x)$ is non-zero, are transcendental numbers as long as $f(x)$ is not a rational multiple of $g(x)$. Indeed, by Theorem 1.10, all the elements of $\mathbb{Q}(\pi)$ that are not rational are transcendental numbers.

We conclude this subsection with another way of producing a variety of transcendental numbers. This construction differs from the ones given in Chapter 18 as, given a transcendental element α over a field L , it allows us

to construct transcendental elements outside the extension $L(\alpha)$. We start by establishing some notation.

Let F be a field and suppose that $\alpha \in F$. Then $\sqrt[n]{\alpha^m}$ denotes any root of the polynomial $x^n - \alpha^m$. When working in $\mathbb{C}$ this notation is normally reserved for the principal root, namely the one with smallest non-negative argument. We are not doing this here, since in general we do not have a definition of ‘principal’ root. This means that the expression $\sqrt[n]{\alpha^m}$ is multi-valued.

Theorem 1.11

Let α be a transcendental element over a field F . Then, for any non-zero relatively prime integers m and n , the element $\sqrt[n]{\alpha^m}$ is transcendental over F .

Proof Let $\beta = \alpha^m$. Then $\beta \in F(\alpha)$. We first show that $\beta = \alpha^m$ is transcendental over F .

Suppose that β is algebraic over F . Then, by Theorem 1.2, $[F(\beta) : F]$ is finite, say equal to n . If m is positive, then $[F(\alpha) : F(\beta)] = m$. Then, by the KLM Theorem (Chapter 18, Theorem 2.29), we have $[F(\alpha) : F] = mn$, which implies that α is algebraic over F , a contradiction. If m is negative, then $\alpha = \beta^{-m} \in F(\beta)$ and so α is not transcendental, a contradiction again. Thus β is transcendental over F .

The element $\sqrt[n]{\beta}$ represents any solution of the equation $x^n = \beta$ and so, by Theorem 4.2 of Chapter 18, it is an element of the splitting field K for the polynomial $x^n - \beta$ over the field $F(\alpha)$.

Suppose that $\sqrt[n]{\beta}$ is algebraic over F . Then, by Theorem 1.5, $(\sqrt[n]{\beta})^n = \beta$ would also be algebraic, contrary to what we have just proved about β . It follows that $\sqrt[n]{\alpha^m}$ is transcendental over F . ■

As a consequence, all the non-zero rational powers of both e and π , for example numbers such as $\sqrt[19]{\pi^{16}}$, are transcendental over $\mathbb{Q}$.

Exercise 1.6

For each of the following elements, decide whether it is algebraic or transcendental over the given field. Briefly justify your decision.

- (a) $\sqrt{e}$ over $\mathbb{Q}$
- (b) $\sqrt{e}$ over $\mathbb{Q}(e)$
- (c) $e^{i\pi}$ over $\mathbb{Q}$

We conclude this section with a summary of the various results we have obtained about field extensions.

Summary – Field extensions

Let E be a field with a subfield F .

- An element $\alpha \in E$ is said to be algebraic over F if α is a root of some polynomial in $F[x]$. An element $\alpha \in E$ is called transcendental over F if α is not a root of any non-zero polynomial in $F[x]$. (Definition 2.11 of Chapter 18)
- If $\alpha \in E \setminus F$ is algebraic over F then
 - α has a minimal polynomial $m(x)$ with $m(\alpha) = 0$
 - $F(\alpha) \cong F[x]/\langle m(x) \rangle$
 - $[F(\alpha) : F] = \deg(m)$
 - $F(\alpha) = \{a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1} : a_0, a_1, \dots, a_{n-1} \in F\}$ where $n = \deg(m)$ and addition and multiplication of elements are carried out modulo $m(x)$.

(Theorem 2.21, Theorem 2.19 and Definition 2.22 of Chapter 18)

- Let K, L and M be fields such that $K \subset L \subset M$ and let the degrees $[M : L]$ and $[L : K]$ be both finite. Then, $[M : L][L : K] = [M : K]$. (Theorem 2.29 of Chapter 18)
- An element α of E is algebraic over F if, and only if, the degree $[F(\alpha) : F]$ is finite. (Theorem 1.2)
- The degree $[E : F]$ is finite if, and only if, E is an algebraic extension of F and there exists a finite subset S of E such that $E = F(S)$. (Theorem 1.3)
- If $h(x)$ is an irreducible polynomial over a field F and α and β are roots of $h(x)$, then $F(\alpha) \cong F(\beta)$. (Exercise 2.5 of Chapter 18)
- If $f(x)$ is a polynomial over F , then up to isomorphism there is a unique extension of F – the splitting field of f – that contains all the roots of f . Moreover, every extension of F over which f splits contains a subfield that is isomorphic to the splitting field. (Definition 4.1, Proposition 4.4 and Corollary 4.7 of Chapter 18)
- Given a field F , the field of rational functions over F is a transcendental extension of F . (Theorem 2.31 of Chapter 18 and Definition 1.9)
- If α, β are transcendental elements over F , then $F(\alpha) \cong F(\beta)$. (Corollary 2.32 of Chapter 18)
- If α is transcendental over F and $\beta \in F(\alpha)$ but $\beta \notin F$ then β is transcendental over F . (Theorem 1.10)
- If α is transcendental over F and m and n are relatively prime integers then $\sqrt[n]{\alpha^m}$ is transcendental over F . (Theorem 1.11)

2 Ruler and compasses constructions

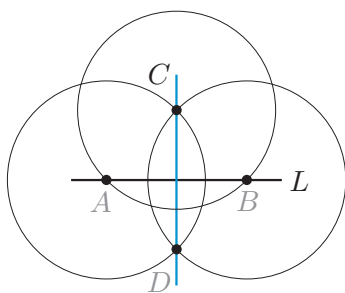
Ruler and compasses constructions are a heritage of the geometry developed by the ancient Greeks. In their view, figures constructed by ruler and compasses were a stamp of perfection. It was this view that was the driving force behind trying to find ways to construct various objects by using just a ruler (with no marks at all) and compasses.

2.1 Geometric constructions

You may be familiar with a number of simple constructions made with the help of ruler and compasses only. Examples include the construction of a line perpendicular to a given line through a given point (see Figure 2.1 and Figure 2.2) and the construction of a 60° angle (see Figure 2.3). You may wish to try out some of these constructions yourself. You can either do this using paper, pencil, ruler and compasses or you can use readily available software on the web or mobile apps that enable you to simulate constructions on screen.

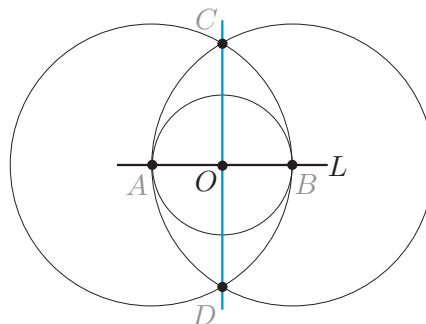
In the examples that follow, we do not list the steps of more basic constructions, such as finding the perpendicular to a given line through a given point, or finding the midpoint of a line segment. You will see these basic constructions in detail in Subsection 2.2.

When we say that we ‘construct’ an angle, we mean that given a line and a point we can construct a line through that point that makes the required angle with the given line.



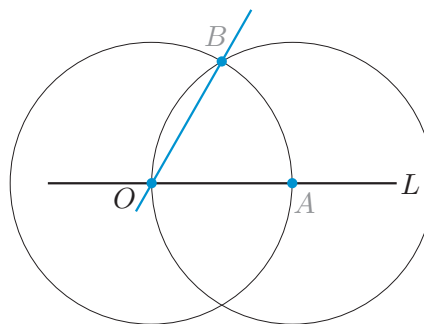
- Given a line L and a point C , draw a circle of centre C cutting L .
- Mark the intersection points with L as A and B .
- Draw two circles with centres A and B and with the same radius as the first circle.
- These circles will intersect at C and another point that we label as D . The line CD passes through C and is perpendicular to the line L .

Figure 2.1 To draw a line perpendicular to a given line L through a given point C



- Given a line L and a point O on L , draw a circle with centre O .
- Mark the intersection points with L as A and B .
- Draw circles, centres A and B , with radius AB .
- Mark the intersection points of these two circles as C and D . CD is a line through O and perpendicular to the line L .

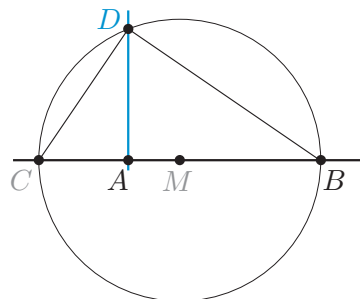
Figure 2.2 To draw a line perpendicular to a given line L through a given point O on L



- Given a line L , mark two distinct points O and A on L .
- Draw a circle of centre O and radius OA .
- Draw a circle of centre A and radius OA .
- These circles will intersect at a point that we label as B . Then the line through O and B is at angle 60° to the line L .

Figure 2.3 To draw an angle of 60°

Another interesting construction is that of the square root of a given length. This is illustrated in Figure 2.4 below, where we assume that we are given both a length 1 and the length whose root is to be determined.



To draw a line whose length is the square root of the given length AB :

- extend the line AB to CB so that $AC = 1$
- find the midpoint M of the line CB (see Exercise 2.1)
- draw the circle of centre M and radius MB
- draw the perpendicular to AB at A , and mark the intersection point with the circle as D
- the length AD is the square root of the length AB , since $\frac{AD}{AC} = \frac{AB}{AD}$ and $(AD)^2 = AC \cdot AB$ so that $(AD)^2 = AB$ as $AC = 1$.

Figure 2.4 Constructing a square root

We mention another famous example without illustrating it: given a construction for a square of a given side, say of length x , we can draw its diagonal, which has length $\sqrt{2}x$. Then the construction can be applied to the diagonal, giving a square whose area is twice the area of the original square.

A construction of a similar flavour is that of a square whose area is equal to the area of a given rectangle. This construction is shown in Figure 2.5 and the steps are given below.

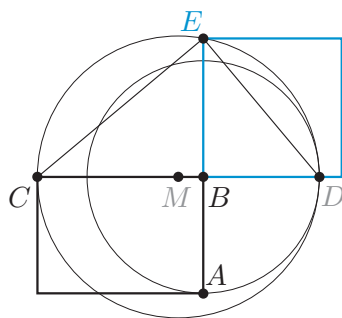


Figure 2.5 Constructing a square with the same area as a given rectangle

Start with the rectangle with vertices at A, B and C .

- Extend the line CB to the right.
- Draw a circle of centre B and radius AB .
- Mark the intersection of the circle with CB , to the right of B , as D .
- Find the midpoint M of the line CD .
- Draw a circle of centre M and radius MD .
- Mark the intersection of this circle with the extended line AB as E .
- Draw a square with side BE .
- The area of the square is the same as that of the rectangle.
- The triangle CBE is similar to the triangle EBD , so $\frac{BC}{EB} = \frac{EB}{BD} = \frac{EB}{AB}$ and therefore $(EB)^2 = BC \cdot AB$.

A large number of other interesting examples exist. There are, however, a number of problems in plane and space geometry that cannot be solved by ruler and compasses only. The three famous problems of this kind which ancient Greeks tried to solve by ruler and compasses but without success were the duplication of a cube, the trisection of an angle, and the quadrature of a circle. The problems ask for a ruler and compasses construction of the side of a cube twice the volume of a given cube, an angle one-third the size of a given angle, and the side of a square of area equal to the area of a given circle, respectively.

You may have come across these three classical problems before because they frequently appear in articles and programs about mathematics for the general public. Our aim in this section is to give rigorous proofs of the impossibility of solving them by ruler and compasses only. The key is to couple geometry with algebra by first using coordinates to translate the geometric nature of the problems into algebraic terms and then apply the theory of field extensions developed in Chapter 18 and in this chapter. This, of course, was not known to ancient Greeks, who apparently did not even have suspicions about possible non-existence of solutions.

2.2 Ruler and compasses algebraically

This subsection develops an algebraic description of the rather intuitive concept of a construction by means of ruler and compasses only. You might have observed that in all the constructions discussed in Subsection 2.1, including the ‘impossible’ ones, we assume that an initial set of points is given, and out of this ‘input set’ we construct, using ruler and compasses only, an ‘output set’ of points that, in one way or another, tells us what the solution is.

A formal description of constructions by ruler and compasses may thus be given as follows.

Definition 2.1 Constructible sets of points

Let S be a set of points in the Euclidean plane that contains at least two points. We introduce two operations corresponding to performing an elementary ‘move’ using ruler and compasses:

O1 Ruler operation Draw a straight line through any two points of S .

O2 Compasses operation Draw a circle whose centre is a point of S and whose radius is equal to the distance between some pair of points in S .

The points of intersection of any two distinct lines or circles drawn using operations O1 or O2 are said to be **constructible from a set S** in one move.

We say that a point B is **constructible** from an initial set S_0 of points in the Euclidean plane if there is a finite sequence $A_1, A_2, \dots, A_k = B$ of points in the plane such that, for $i = 1, 2, \dots, k$, the point A_i is constructible in one move from the set $S_0 \cup \{A_1, \dots, A_{i-1}\}$.

We illustrate this formal definition of constructibility on the example of finding the centroid of a triangle ABC , that is, the point of intersection of the three medians of the triangle. The idea is to bisect the segments AB and BC ; the centroid then lies on the intersection of the line through C and the midpoint of the segment AB and the line through A and the midpoint of BC , as shown in Figure 2.6.

Remember that a median is a line connecting one vertex of the triangle with the midpoint of the opposite side.

The construction lines for the midpoints have been omitted.

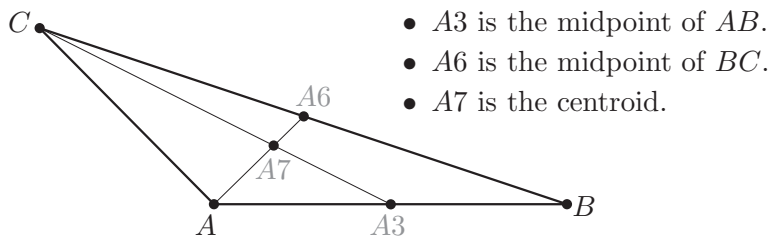


Figure 2.6 Constructing the centroid of a triangle

Example 2.2 *The centroid of a triangle*

We may assume that we are given the three vertices A , B and C of a triangle. Thus, our ‘input set’ will be $S_0 = \{A, B, C\}$.

1. Draw the circle centred at A of radius AB . (O2)
2. Draw the circle centred at B of radius AB . (O2)
3. Let A_1 and A_2 be the points of intersection of the two circles.
4. Draw the line through A_1 and A_2 . (O1)
5. Let A_3 be the point of intersection of the lines through A_1 , A_2 and through A , B .
6. Draw the circle centred at B of radius BC . (O2)
7. Draw the circle centred at C of radius BC . (O2)
8. Let A_4 and A_5 be the points of intersection of the two circles.
9. Draw the line through A_4 and A_5 . (O1)
10. Let A_6 be the point of intersection of the lines through A_4 , A_5 and through B , C .
11. Draw the line through C and A_3 . (O1)
12. Draw the line through A and A_6 . (O1)
13. Let A_7 be the point of intersection of the last two lines. This point is the centroid of the triangle.

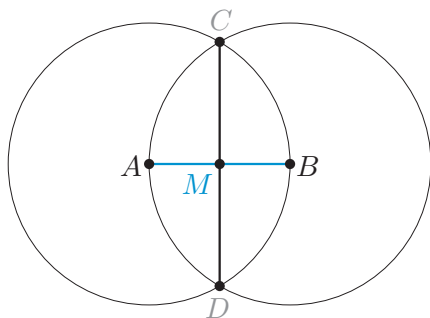
In our terminology, the sequence $A_1, A_2, A_3, A_4, A_5, A_6, A_7$ defines a construction of the centroid of the triangle ABC from the input set $\{A, B, C\}$. Figure 2.6 shows some of the details of this construction.

Our formal definition of constructibility indeed captures all the ruler and compasses constructions we have met in Euclidean geometry. This is because to draw a line requires two specified points on the line, and to draw a circle requires having its centre and radius, these being again given by two specified points (the centre and a point on the circumference).

Exercise 2.1

Describe the process of drawing the midpoint of a segment given its two endpoints in the same way as in Example 2.2, that is, label each operation in the process as O1 or O2. Figure 2.7 shows the standard construction for a midpoint.

For this exercise you should assume that the compass collapses when lifted off the page and so cannot be used to ‘transfer’ distance.

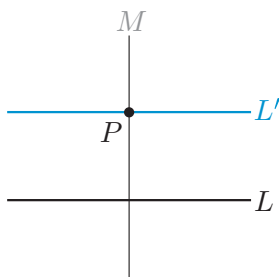


- With centre A , draw a circle of radius AB .
- With centre B , draw a circle of radius AB .
- Label the two intersections of the circles as C and D .
- The line CD is the perpendicular bisector to AB .
- The midpoint M is the intersection of CD and AB .

Figure 2.7 Finding the midpoint of a segment of given endpoints

Exercise 2.2

Describe the process of drawing a line parallel to another line L through a given point P in the same way as in Example 2.2, that is, label each operation as O1 or O2. Figure 2.8 gives a standard construction that involves drawing a perpendicular, M , to L through P . The line through P and perpendicular to M is then the required parallel.



- The line L and the point P are given.
- Draw the perpendicular, M , from P to L .
- Draw the perpendicular to M at P .
- Then L' is the line through P , parallel to L .

Figure 2.8 To draw a line through a point P parallel to another line

Exercise 2.3

Describe how to draw a circle with centre C and radius equal to AB given three points A , B and C not all lying on the same line. You may use the construction of a parallel line described in Exercise 2.2.

The solution to Exercise 2.3 shows that we can

- copy a length, that is, given a line segment AB and point C on a line L we can construct a point D on L so that AB and CD have the same length;
- always draw a circle of radius 1, since an input set S will always contain two points and these can be labelled as $(0,0), (1,0)$.

Exercise 2.4

Given two lines meeting at a point V and another line with a point A , describe how to draw a line through A so that the angle made by the lines meeting at A is the same as the angle between the two lines meeting at V (you may assume that you can copy a length and draw a circle of radius 1).

At this point it is worth collecting some standard constructions that we can incorporate into descriptions without giving full details.

C1 Copy a length.

C2 Draw a circle of radius 1.

C3 Draw a line through the origin at 60° (see Figure 2.3).

C4 Draw a perpendicular to a line through a given point.

C5 Find the midpoint of two points.

C6 Draw a line parallel to a given line through a point not on the given line.

C7 Copy an angle.

Exercise 2.5

Describe the following constructions using the operations O1 and O2 and the constructions C1–C7.

- Given a point $(a,0)$, construct the points $(\frac{a}{3},0)$ and $(\frac{2a}{3},0)$.
- Divide the line between $(0,0)$ and $(a,0)$ into n equal parts.
- Construct the tangent to a circle from a given point not on the circle.
- Given two circles with different radii and one not inside the other, construct the two common external tangents as shown in Figure 2.9.

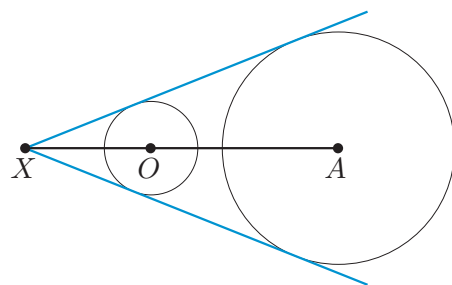


Figure 2.9

Hint: for part (d), use the fact that a tangent is perpendicular to a radius so that the radii that meet the tangent in the two circles are parallel. Then use some similar triangles to locate the point X on the diagram. Finally, use the fact that the angle subtended on the circumference by a diameter is a right angle.

2.3 Coordinates and field extensions

Recall that our formal construction resulted in a sequence $A_1, A_2, \dots, A_k$ of points obtained from an initial set S_0 of points. We now extend this idea to sequences and sets of real numbers.

Definition 2.3 Constructible number

- (a) We say that a real number a is **constructible** if the point $(a, 0)$ is constructible from the set of points $\{(0, 0), (1, 0)\}$.
- (b) Let S be a set of points in the plane that contains at least two points. We first determine a coordinate system. Choose two points in S and let one point be the origin, $(0, 0)$ and the other point be the point $(1, 0)$. We say that a real number a is **constructible from a set S** , with the given coordinate system, if the point $(a, 0)$ is constructible from the set of points S .

We first show that a set of constructible numbers forms a field.

Proposition 2.4

The set of constructible numbers from a set S containing at least two points forms a subfield of $\mathbb{R}$.

Proof We prove the theorem by showing that if a, b are non-zero constructible numbers from a set S , then the numbers $-a$, $a + b$, ab and $\frac{a}{b}$ are constructible as well. We first note that our definition of constructible points and numbers tells us that the points $O = (0, 0)$ and $I = (1, 0)$ (and hence the numbers 0 and 1) are constructible. The line through these points is the x -axis.

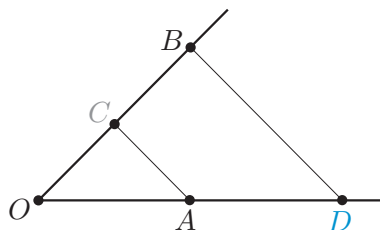
Let a, b be constructible, so that the points $A = (a, 0)$ and $B = (b, 0)$ are constructible. Firstly, draw a circle centred at O through A . The point $(-a, 0)$ is the intersection of this circle with the line AB that is not A , which implies that $-a$ is constructible.

To demonstrate the constructibility of $a + b$ we assume without loss of generality that $b > a > 0$. The construction is as follows.

1. Draw the circle centred at B with radius a . (O2)
2. This circle intersects the line through O and B in two points; mark the one lying outside the segment OB as $C = (a + b, 0)$.

This shows that $(a + b, 0)$ is constructible.

Figure 2.10 shows the construction of $(ab, 0)$ for $a, b > 0$. The construction lines for the parallel lines have been omitted.



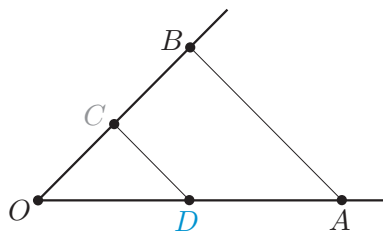
- OA has length a .
- OC has length 1 and OB has length b .
- D is constructed so that BD is parallel to CA .
- Triangle OAC is similar to triangle ODB .
- $\frac{OA}{OC} = \frac{OD}{OB}$ and so $OD = ab$.

Figure 2.10 Constructing ab from a and b

1. Draw the line through the origin O and $A = (a, 0)$. This is the x -axis. (O1)
2. Draw a line, L , through the origin, that does not pass through $I (= (1, 0))$. (O1)
3. Mark points C and B at distance 1 and b from the origin respectively. (C1)
4. Draw the line through $A (= (a, 0))$ and C . (O1)
5. Construct a line through B that is parallel to the line AC . (C6)
6. Mark the intersection of this line with the x -axis as D .

Since triangles OAC and ODB are similar, we have that $\frac{OA}{OC} = \frac{OD}{OB}$ and so $OD = ab$, that is, D is the point $(ab, 0)$.

The construction of $(\frac{a}{b}, 0)$ for $a, b > 0$ is shown in Figure 2.11. As in Figure 2.10, the construction lines for parallel lines have been omitted.



- OA has length a .
- OC has length 1 and OB has length b .
- D is constructed so that CD is parallel to BA .
- Triangle ODC is similar to triangle OAB .
- $\frac{OD}{OC} = \frac{OA}{OB}$ and so $OD = \frac{a}{b}$.

Figure 2.11 Constructing $\frac{a}{b}$ from a and b

1. Draw the line through the origin O and $A (= (a, 0))$. This is the x -axis. (O1)
2. Draw a line, L , through the origin, that does not pass through $I (= (1, 0))$. (O1)
3. Mark points C and B at distance 1 and b from the origin respectively. (C1)
4. Draw the line through A and B . (O1)
5. Construct a line through C that is parallel to the line AB . (C6)
6. Mark the intersection of this line with the x -axis as D .

Since triangles ODC and OAB are similar we have that $\frac{OD}{OC} = \frac{OA}{OB}$ and so $OD = \frac{a}{b}$, that is D is the point $(\frac{a}{b}, 0)$.

We have shown that additive inverses, sums, products and quotients of constructible numbers are constructible. This implies that the set of all the numbers constructible from a given set S containing at least two points is a subfield of $\mathbb{R}$, as claimed. ■

We have shown that the set of constructible points formed from a starting set S is a subfield, K say, of $\mathbb{R}$. We now look at certain subfields of K obtained by adjoining the coordinates of a point constructed from a given set of points in one move.

Lemma 2.5

Let S be a set of at least two points in the plane and let K be the subfield of $\mathbb{R}$ generated by the x - and y -coordinates of the points in S . Let $A = (x, y)$ be a point constructed in one move from the set S and let $L = K(x, y)$ be the subfield of $\mathbb{R}$ obtained by adjoining x and y to K . Then the degree $[L : K]$ is either 1 or 2.

Proof The claim will be verified if we show that x and y are real roots of linear or quadratic polynomials with coefficients in K .

We have to consider three cases, since the point A can arise as an intersection of two lines, a line and a circle, and two non-concentric circles. In each case we make use of the fact that a line in $\mathbb{R}^2$ can be identified with the set of all pairs $(x, y) \in \mathbb{R}^2$ that satisfy a linear equation $ax + by = c$ for some real a, b, c such that $(a, b) \neq (0, 0)$, and a circle can be identified with the set of all pairs $(x, y) \in \mathbb{R}^2$ that satisfy a quadratic equation of the form $(x - s)^2 + (y - t)^2 = r^2$ for some real r, s, t such that $r > 0$.

Suppose first that the point A has been obtained as an intersection of two distinct lines, with equations, say, $a_1x + b_1y = c_1$ and $a_2x + b_2y = c_2$. By our constructibility assumption, both lines have been constructed from points in S , that is, we assume that all the six coefficients belong to K . From elementary algebra we know that the solution is

$$x = \frac{c_1b_2 - c_2b_1}{a_1b_2 - a_2b_1} \quad \text{and} \quad y = \frac{a_1c_2 - a_2c_1}{a_1b_2 - a_2b_1},$$

that is, $x, y \in K$.

Secondly, let A arise as a point of intersection of a line given by the equation $ax + by = c$ with a circle given by $(x - s)^2 + (y - t)^2 = r^2$. By the same token as above, we assume that all the five coefficients a, b, c, s, t are in K . We do not know whether r is an element of K , but since r^2 is a distance of two points with coordinates in K , we definitely know that $r^2 \in K$. We know that $(a, b) \neq (0, 0)$; without loss of generality, assume that $b \neq 0$. Then, out of the equation of the line we have $y = \frac{c - ax}{b}$. Substituting this into the equation of the circle gives

$$(x - s)^2 + \left(\frac{c - ax}{b} - t \right)^2 = r^2,$$

which is a quadratic equation with coefficients in K . Therefore, the coordinates x and y in this case are roots of quadratic polynomials over K and hence $[L : K] \leq 2$, as claimed.

The third and final case to consider is when A arises as a point of intersection of two circles with distinct centres, with equations, say, $(x - s_1)^2 + (y - t_1)^2 = r_1^2$ and $(x - s_2)^2 + (y - t_2)^2 = r_2^2$. Again, we know that $s_1, t_1, s_2, t_2 \in K$ and $r_1^2, r_2^2 \in K$. Since the circles cannot have the same centre it is not possible that $s_1 = s_2$ and $t_1 = t_2$. Expanding the brackets in the two equations gives

$$\begin{aligned} x^2 + y^2 - 2s_1x - 2t_1y &= r_1^2 - s_1^2 - t_1^2, \\ x^2 + y^2 - 2s_2x - 2t_2y &= r_2^2 - s_2^2 - t_2^2. \end{aligned}$$

Subtracting the second equation from the first yields

$$2(s_2 - s_1)x + 2(t_2 - t_1)y = r_1^2 - r_2^2 + s_2^2 - s_1^2 + t_2^2 - t_1^2.$$

Since $s_1 \neq s_2$ or $t_1 \neq t_2$ we can assume, without loss of generality, that $t_1 \neq t_2$. Then we can make y the subject of the third equation and substitute the result into the first. This gives a single quadratic equation for x , with all the coefficients in K . The unknown y is then evaluated from the third equation in terms of x .

Summing up, we have proved that $L = K(x, y)$ arises from K by adjoining real numbers x and y that are roots of linear or quadratic polynomials over K . The result follows. ■

We now derive our main result, a necessary condition for constructibility.

Theorem 2.6

Suppose that the point $A = (x, y)$ is constructible from a set S of at least two points in $\mathbb{R}^2$. Let E be the subfield of $\mathbb{R}$ generated by the coordinates of the points in S . Then the degree $[E(x, y) : E]$ is a power of 2.

Proof Let $A_1, \dots, A_k = A$ be the finite sequence of points in the plane such that, for $i = 1, 2, \dots, k$, the point $A_i = (x_i, y_i)$ is constructible in one move from the set $S \cup \{A_1, \dots, A_{i-1}\}$; note that $(x_k, y_k) = (x, y)$. Further, let $E_0 = E$, and for $i = 1, 2, \dots, k$ let E_{i-1} be the subfield generated by the x - and y -coordinates of the points in $S \cup \{A_1, \dots, A_{i-1}\}$.

By Lemma 2.5 applied to the fields $K = E_{i-1}$ and $L = E_i = E_{i-1}(x_i, y_i)$ we see that the degree $[E_i : E_{i-1}]$ is equal to 1 or 2. The KLM Theorem (Chapter 18, Theorem 2.29) then gives

$$[E_k : E_0] = [E_{k-1}(x_k, y_k) : E_{k-1}] \cdots [E_{i-1}(x_i, y_i) : E_{i-1}] \cdots [E_0(x_1, y_1) : E_0]$$

where each term in the product on the right-hand side is 1 or 2. It follows that $[E_k : E_0]$ is a power of 2.

Now $E(x, y) = E_0(x, y)$ is an extension of $E = E_0$ and a subfield of E_k . Thus, $[E(x, y) : E]$ must divide $[E_k : E_0]$ and hence $[E(x, y) : E]$ must be a power of 2. ■

This result will mostly be used in its contrapositive form, which we state next.

Theorem 2.7

Let S be a set of points in the real plane $\mathbb{R}^2$ and let E be the subfield of $\mathbb{R}$ generated by the coordinates of all the points in S . Suppose that $A = (x, y)$ is a point in $\mathbb{R}^2$ such that the degree of the extension $[E(x, y) : E]$ is not a power of 2. Then the point A is not constructible from S .

We can then use Theorem 2.21(c) of Chapter 18 to state this result in terms of minimal polynomials; for our purposes, it will be sufficient to assume that the point in question lies on the x -axis.

Corollary 2.8

Let S be a set of points in the real plane $\mathbb{R}^2$ and let E be the subfield of $\mathbb{R}$ generated by the coordinates of all the points in S . Let $A = (x, 0)$ be a point in $\mathbb{R}^2$ such that the minimal polynomial for x over the field E has a degree that is not a power of 2. Then x is not a constructible number.

In the next subsection we will apply this result to prove the impossibility of constructing a solution for each of the three classical problems the ancient Greeks could not solve.

Exercise 2.6

Let $S = \{(0, 0), (1, 0)\}$. Why is every point (a, b) with $a, b \in \mathbb{Q}$ constructible?

Exercise 2.7

- (a) Show that the perpendicular bisector of the segment with endpoints $A = (1, 2)$ and $B = (3, 4)$ is the line with equation $x + y - 5 = 0$.
- (b) Use the equation of the perpendicular bisector found in (a) to find the point of intersection of the line AB and this perpendicular bisector, that is, the midpoint of AB .

2.4 Impossibility proofs for classical Greek problems

We are about to prove that there do not exist any ruler and compasses constructions for three classical problems of Greek geometry: duplication of a cube, trisection of an angle, and quadrature of a circle. Duplicating the cube means constructing a cube of twice the volume of a given cube, and quadrature of the circle is another way of saying squaring the circle, namely, constructing a square of the same area as a given circle.

Theorem 2.9

The cube cannot be duplicated using ruler and compasses only.

Proof Assume that the side of a cube to be duplicated is given. We may assume that the side is placed in the real plane $\mathbb{R}^2$ in such a way that the segment representing it has endpoints with coordinates $A = (0, 0)$ and $B = (1, 0)$ – as this representation only depends on the choice of units on the coordinate axes, our assumption causes no loss of generality. We may therefore assume that our starting set is $\{A, B\}$. Then the smallest subfield of $\mathbb{R}$ containing the coordinates of points in S is $\mathbb{Q}$.

If it were possible to duplicate the cube with ruler and compasses, we would be able to construct the point $(b, 0)$ where $b^3 = 2$, that is, $b = \sqrt[3]{2}$. But as we saw in Example 2.26 of Chapter 18, $[\mathbb{Q}(b) : \mathbb{Q}] = 3$, which is not a power of 2. By Theorem 2.7, such a point is not constructible, and hence duplication of the cube is impossible. ■

Theorem 2.10

The angle $\pi/3$ cannot be trisected by a construction using ruler and compasses only.

Proof We may assume without loss of generality that the angle is given by its vertex $A = (0, 0)$ and two other points $B = (1, 0)$ on the x -axis and $C = (\cos(\pi/3), \sin(\pi/3)) = (\frac{1}{2}, \frac{\sqrt{3}}{2})$ so that the points B, C lie on the unit circle centred at A and BAC is our angle of magnitude $\pi/3$. Note that the point C is actually constructible from A and B and is therefore not necessary (see Figure 2.3).

To trisect the angle $\pi/3$ is then equivalent to constructing the point $D = (d, 0)$ where $d = \cos(\pi/9)$. If D were constructible, then, by doubling the segment AD , we could also construct the point $\Delta = (\delta, 0)$ such that $\delta = 2\cos(\pi/9)$.

At this point we need a formula from trigonometry that can be derived from the formulas for $\cos(\alpha + \beta)$ and $\sin(2\alpha)$:

$$\cos(3\alpha) = 4\cos^3\alpha - 3\cos\alpha.$$

Now, if $\alpha = \pi/9$, we have $\cos(3\alpha) = \frac{1}{2}$, so that our equation multiplied by 2 gives $1 = (2\cos(\pi/9))^3 - 3(2\cos(\pi/9))$. In other words, $\delta = 2\cos(\pi/9)$ is a root of the polynomial $x^3 - 3x - 1$.

By Exercise 2.4 of Chapter 18, however, if α is a root of the polynomial $x^3 - 3x - 1$, then $[\mathbb{Q}(\alpha) : \mathbb{Q}] = 3$, which is not a power of 2. Theorem 2.7 then tells us that the point Δ is not constructible and therefore the point D is not constructible. This shows the impossibility of trisection of the angle $\pi/3$. ■

Theorem 2.11

The quadrature of a circle cannot be constructed by using ruler and compasses only.

Proof By scaling we may assume that the circle to be ‘squared’ is the unit circle centred at the origin. The area of this circle is π . Therefore, finding the quadrature of this circle means constructing the point $(\sqrt{\pi}, 0)$ from the origin and the point $(1, 0)$. If this were possible, then the degree $[\mathbb{Q}(\sqrt{\pi}) : \mathbb{Q}]$ would have to be a power of 2. In particular, $\sqrt{\pi}$ and hence π , would have to be algebraic numbers. But a famous theorem by Lindemann, which is beyond the scope of this course, tells us that the number π is not algebraic. This contradiction proves the impossibility of the quadrature of a circle. ■

Exercise 2.8

Can the angle $2\pi/5$ be trisected using ruler and compasses?

Exercise 2.9

Prove that the angle α can be trisected by ruler and compasses if, and only if, the polynomial

$$4x^3 - 3x - \cos(\alpha)$$

is reducible over $\mathbb{Q}(\cos(\alpha))$.

2.5 Constructible complex numbers

The idea of a real constructible number can be extended to a complex number by saying that $a + ib$ is constructible if, and only if, (a, b) is constructible. As i is in a quadratic extension of any subfield of the real numbers (since it is a root of the quadratic polynomial $x^2 + 1$), we can immediately extend our main constructibility result to the complex case.

Theorem 2.12

Let S be a set of complex numbers (including 1 and 0) and let E be the subfield of $\mathbb{C}$ generated by S . Suppose that $a + ib$ is a complex number such that the degree of the extension $[E(a + ib) : E]$ is not a power of 2. Then $a + ib$ is not constructible from S .

We can interpret this theorem in terms of the minimal polynomial of a complex number just as we did for real numbers in Corollary 2.8.

Corollary 2.13

Let K be the field of constructible complex numbers. If the degree of the minimal polynomial for the complex number $a + ib$ over $\mathbb{Q}$ is not a power of 2, then $a + ib \notin K$.

With this result to hand we can combine this with the work on cyclotomic polynomials discussed just before Lemma 4.17 of Book C, Chapter 11, to show that certain roots of unity are not constructible.

Proposition 2.14

Let p be a prime number and let x be a root of $x^p - 1$ with $x \neq 1$. If x is constructible, then p is of the form $2^n + 1$.

Proof Now $x^p - 1 = (x - 1)(x^{p-1} + x^{p-2} + \cdots + x + 1)$, and in Book C, Chapter 11, Worked Exercise 4.18 we showed that $x^{p-1} + x^{p-2} + \cdots + x + 1$ is irreducible when p is prime. Thus $x^{p-1} + x^{p-2} + \cdots + x + 1$ is the minimal polynomial for x , and so if $p - 1$ is not a power of 2, x is not constructible. The result then follows. ■

If we can construct a regular polygon with n sides, we can construct an n th root of unity by making the ‘radius’ of our polygon 1 and positioning a vertex at $(1, 0)$. Hence if a regular polygon has p sides, where p is a prime number not of the form $2^n + 1$, then it is not constructible. We can see immediately that it is not possible to construct regular polygons with 7, 11, or 13 sides.

Exercise 2.10

Show that it is impossible to construct a regular 9-gon using ruler and compasses.

Exercise 2.11

This exercise asks you to find two constructions of the regular pentagon.

- (a) Let $c = 2 \cos\left(\frac{2\pi}{5}\right)$ and let $x = \exp\left(\frac{2i\pi}{5}\right)$, that is, x is a fifth root of unity, so that $x^4 + x^3 + x^2 + x + 1 = 0$. Show that $c = x + \frac{1}{x}$, and hence show that $c^2 + c - 1 = 0$. Deduce that $\cos\left(\frac{2\pi}{5}\right) = \frac{-1+\sqrt{5}}{4}$. Use this to find a construction for the regular pentagon.
- (b) This construction is given by Ptolemy in *The Almagest*.
 - (i) Show that $\cos\left(\frac{\pi}{5}\right) = \frac{1+\sqrt{5}}{4}$.
 - (ii) Use the cosine rule to show that the length of the side of a decagon, inscribed in a circle of radius 1, has length $\frac{\sqrt{5}-1}{2}$.
 - (iii) Construct a regular pentagon.

Ptolemy, translated by R. C. Taliaferro, *Almagest I–IV* ‘On the size of chords in a circle’, pp. 14–15.

Learning outcomes

We list those topics on which we may set assessment questions to test your understanding of the chapter.

After working through this chapter you should be able to:

- (a) Find a polynomial equation for an algebraic element over a field in simple cases.
- (b) Determine whether a given element of a field extension is transcendental or algebraic in simple cases.
- (c) Describe a simple ruler and compasses construction in terms of basic constructions.
- (d) Determine whether a given number, point or angle is constructible from a given initial set of points.
- (e) Prove simple results about field extensions and ruler and compasses constructions.

Solutions to exercises

Solution to Exercise 1.1

- (a) We use the subfield criterion (Book C, Chapter 11, Lemma 1.21) to show that L is a subfield of E . Let $a, b \in L$. If M_i is a subfield of E that contains both S and F then $a, b \in M_i$ and so $a - b \in M_i, ab \in M_i$ and if $b \neq 0$ then $ab^{-1} \in M_i$. Since this is true for all such fields M_i we have $a - b \in L$ and if $b \neq 0$ then $ab^{-1} \in L$. Thus L is a subfield of E .
- (b) It is clear that L contains both S and F . If K is another subfield of E that contains both S and F , then L will be a subfield of K since $L = \bigcap M_i$ and $K = M_i$ for some i .

Solution to Exercise 1.2

Since α is algebraic over F there is a polynomial $f(x)$ over F with root $\alpha \in E$. The polynomial $f(x)$ is also a polynomial over K and so α is algebraic over K as required.

Solution to Exercise 1.3

- (a) Multiplying numerator and denominator by $1 + \sqrt{2} - (2 + \sqrt{2})\sqrt{3}$ gives the required result.
- (b) Since $-15 - 10\sqrt{2} = -5(3 + 2\sqrt{2})$, multiplying the numerator and denominator of $\frac{1 + \sqrt{2} - 2\sqrt{3} - \sqrt{6}}{-15 - 10\sqrt{2}}$ by $3 - 2\sqrt{2}$ gives

$$\frac{1}{1 + \sqrt{2} + 2\sqrt{3} + \sqrt{6}} = \frac{1}{5}(1 - \sqrt{2} + 2\sqrt{3} - \sqrt{6}).$$

Solution to Exercise 1.4

- (a) From the solution to Worked Exercise 1.4 we know that $\sqrt{2} + \sqrt{3}$ is a root. Since only even powers of x appear in this polynomial, $-\sqrt{2} - \sqrt{3}$ is another root. Again from Worked Exercise 1.4 we have that

$$x^4 - 10x^2 + 1 = (x^2 - (5 + 2\sqrt{6}))(x^2 - (5 - 2\sqrt{6}))$$

and so the other two roots are among the square roots of $5 \pm 2\sqrt{6}$.

These square roots can be found by writing out

$(a_0 + a_1\sqrt{2} + a_2\sqrt{3} + a_3\sqrt{6})^2$ and comparing coefficients. In our particular case we can bypass the calculation by observing that $(\sqrt{3} - \sqrt{2})^2 = 5 - 2\sqrt{6}$ and so the required roots are $\pm(\sqrt{3} - \sqrt{2})$.

- (b) Since $\sqrt{2} + \sqrt{3} \in \mathbb{Q}(\sqrt{2}, \sqrt{3})$ we know that $\mathbb{Q}(\sqrt{2} + \sqrt{3}) \subseteq \mathbb{Q}(\sqrt{2}, \sqrt{3})$. In Exercise 2.10 of Chapter 18 you saw that $[\mathbb{Q}(\sqrt{2}, \sqrt{3}) : \mathbb{Q}] = 4$. We have just shown that $\sqrt{2} + \sqrt{3}$ is a root of an irreducible polynomial of degree 4 over $\mathbb{Q}$, and so $[\mathbb{Q}(\sqrt{2} + \sqrt{3}) : \mathbb{Q}] = 4$. Hence $\mathbb{Q}(\sqrt{2}, \sqrt{3}) = \mathbb{Q}(\sqrt{2} + \sqrt{3})$.
- (c) Since $\sqrt{2} + \sqrt{3}$ is a root of $x^4 - 10x^2 + 1$, the extension $\mathbb{Q}(\sqrt{2} + \sqrt{3})$ is contained in the splitting field for $x^4 - 10x^2 + 1$.

Now $(\sqrt{2} + \sqrt{3})^2 = 5 + 2\sqrt{6}$, and so $\sqrt{6} \in \mathbb{Q}(\sqrt{2} + \sqrt{3})$. This implies that $\pm(\sqrt{6} - 1) \in \mathbb{Q}(\sqrt{2} + \sqrt{3})$. Thus $x^4 - 10x^2 + 1$ splits over $\mathbb{Q}(\sqrt{2} + \sqrt{3})$ and so $\mathbb{Q}(\sqrt{2} + \sqrt{3})$ is the splitting field for this polynomial.

Solution to Exercise 1.5

(a) In this instance a direct calculation is easiest.

$$\begin{aligned}(1 + \sqrt{2})^2 &= 1 + 2\sqrt{2} + 2 \\ &= 2(1 + \sqrt{2}) + 1.\end{aligned}$$

Hence a polynomial equation satisfied by $1 + \sqrt{2}$ in the field $\mathbb{Q}(\sqrt{2})$ is $x^2 - 2x - 1 = 0$.

(b) Here we use that approach illustrated in Worked Exercise 1.6. A basis for $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ is

$$u_1 = 1, u_2 = \sqrt{2}, u_3 = \sqrt{3}, u_4 = \sqrt{6}.$$

If $\alpha = \sqrt{2} - \sqrt{3}$ we have

$$\begin{aligned}u_1\alpha &= \sqrt{2} - \sqrt{3} \\ u_2\alpha &= 2 - \sqrt{6} \\ u_3\alpha &= \sqrt{6} - 3 \\ u_4\alpha &= 2\sqrt{3} - 3\sqrt{2},\end{aligned}$$

so that

$$\begin{aligned}u_1\alpha &= u_2 - u_3 \\ u_2\alpha &= 2u_1 - u_4 \\ u_3\alpha &= -3u_1 + u_4 \\ u_4\alpha &= -3u_2 + 2u_3.\end{aligned}$$

This gives the determinant equation

$$\begin{vmatrix} -\alpha & 1 & -1 & 0 \\ 2 & -\alpha & 0 & -1 \\ -3 & 0 & -\alpha & 1 \\ 0 & -3 & 2 & -\alpha \end{vmatrix} = 0.$$

Swapping columns and multiplying by -1 gives

$$\begin{vmatrix} 1 & 0 & \alpha & -1 \\ 0 & 1 & -2 & \alpha \\ \alpha & -1 & 3 & 0 \\ -2 & \alpha & 0 & 3 \end{vmatrix}.$$

Clearing the first column below the diagonal gives:

$$\begin{vmatrix} 1 & 0 & \alpha & -1 \\ 0 & 1 & -2 & \alpha \\ 0 & -1 & 3 - \alpha^2 & \alpha \\ 0 & \alpha & 2\alpha & 1 \end{vmatrix} \quad \begin{aligned} r_3 &= r_3 - \alpha r_1 \\ r_4 &= r_4 + 2r_1. \end{aligned}$$

Clearing the second column below the diagonal gives

$$\begin{vmatrix} 1 & 0 & \alpha & -1 \\ 0 & 1 & -2 & \alpha \\ 0 & 0 & 1 - \alpha^2 & 2\alpha \\ 0 & 0 & 4\alpha & 1 - \alpha^2 \end{vmatrix} \begin{array}{l} r_3 = r_3 + r_2 \\ r_4 = r_4 - \alpha r_2, \end{array}$$

which is equivalent to

$$\begin{vmatrix} 1 - \alpha^2 & 2\alpha \\ 4\alpha & 1 - \alpha^2 \end{vmatrix} = 0.$$

Evaluating the determinant gives

$$(1 - \alpha^2)^2 - 8\alpha^2 = 0$$

and hence

$$\alpha^4 - 10\alpha^2 + 1 = 0.$$

Thus a polynomial equation satisfied by $\sqrt{2} - \sqrt{3}$ in $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ is $x^4 - 10x^2 + 1 = 0$.

Solution to Exercise 1.6

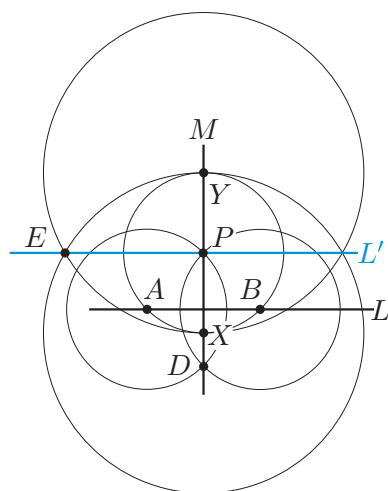
- (a) This is transcendental by Theorem 1.11.
- (b) This is algebraic over $\mathbb{Q}(e)$ since it is a root of the polynomial $x^2 - e$.
- (c) $e^{i\pi} = -1$ and so $e^{i\pi}$ is not transcendental over $\mathbb{Q}$.

Solution to Exercise 2.1

This is in fact the first stage in constructing a centroid so we need the first 5 steps from Example 2.2 with appropriate changes in notation.

1. Draw the circle centred at A of radius AB . (O2)
2. Draw the circle centred at B of radius AB . (O2)
3. Let C and D be the points of intersection of the two circles.
4. Draw the line through C and D . (O1)
5. Let M be the point of intersection of the lines through CD and through AB .

Solution to Exercise 2.2



1. Mark a point A on the line L .
2. Draw a circle C of centre P and radius PA . (O2)
3. Let B be the intersection point of the circle with L that is not A .
4. Draw two circles with centres A and B and radii AP and BP respectively. (O2)
5. Let D be the intersection of the two circles that is not P .
6. Draw the line M through D and P . (O1)
7. Let X and Y be the two intersection points of the circle C with M .
8. Draw two circles with centres X and Y and radius XY . (O2)
9. Let E be an intersection point of the two circles.
10. Draw the line L' through E and P . (O1)

Solution to Exercise 2.3

1. Draw a line parallel to AB through C .
2. Draw a line parallel to AC through B .
3. Mark the intersection of the two lines as D .
4. Draw a circle, centre C and radius CD . (O2)

Since $ABCD$ has been constructed to be a parallelogram, the sides AB and CD will have the same length.

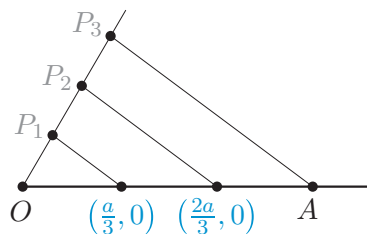
Solution to Exercise 2.4

1. Draw circles of radius 1 with centres at V and A respectively.
2. Mark two adjacent intersections of the lines meeting at V with the circle, centre V , as B and C . (The word adjacent is used here since the lines cross at V and give four intersections with the circle.)
3. Mark an intersection of the line through A with the circle of centre A , as D .
4. Draw a circle of centre D and radius BC .
5. Mark an intersection of this circle with the circle of centre A as E .
6. Draw the line AE . (O1)

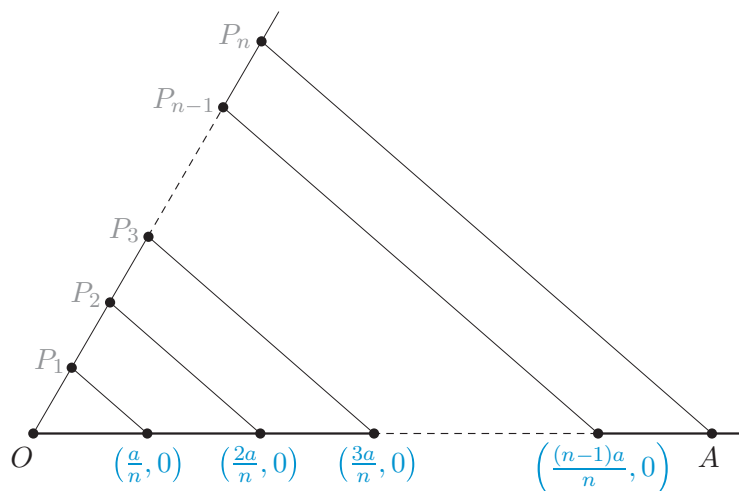
The angle DAE will be the same as the angle BVC .

Solution to Exercise 2.5

- (a) This construction uses the fact that the ratios of corresponding sides of similar triangles are equal.
1. Draw a line through the origin at 60° . (C3)
 2. Mark points P_1, P_2, P_3 along this new line at distance 1, 2, 3 from the origin respectively. (C1)
 3. Draw the line L_3 from P_3 to $(a, 0)$. (O1)
 4. Draw L_1 and L_2 parallel to L_3 through P_1 and P_2 respectively. (C6)
 5. The intersections of L_1 and L_2 with the x -axis will give the points $(\frac{a}{3}, 0)$ and $(\frac{2a}{3}, 0)$ respectively.

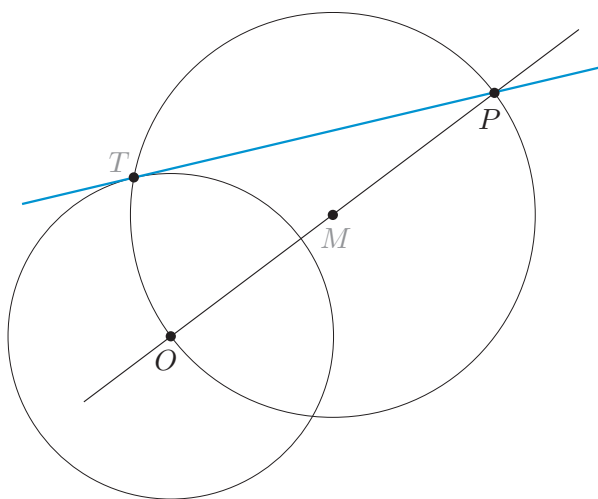


- (b) This construction uses the fact that the ratios of corresponding sides of similar triangles are equal.
1. Draw a line through the origin at 60° . (C3)
 2. Mark points $P_1, P_2, \dots, P_n$ at distance $1, 2, \dots, n$ from the origin respectively. (C1)
 3. Draw the line L_n from P_n to $(a, 0)$. (O1)
 4. Draw $L_1, L_2, \dots, L_{n-1}$ parallel to L_n through $P_1, P_2, \dots, P_{n-1}$ respectively. (C6)
 5. The intersections of $L_1, L_2, \dots, L_{n-1}$ with the x -axis will give the points $(\frac{a}{n}, 0), (\frac{2a}{n}, 0), \dots, (\frac{(n-1)a}{n}, 0)$ respectively.



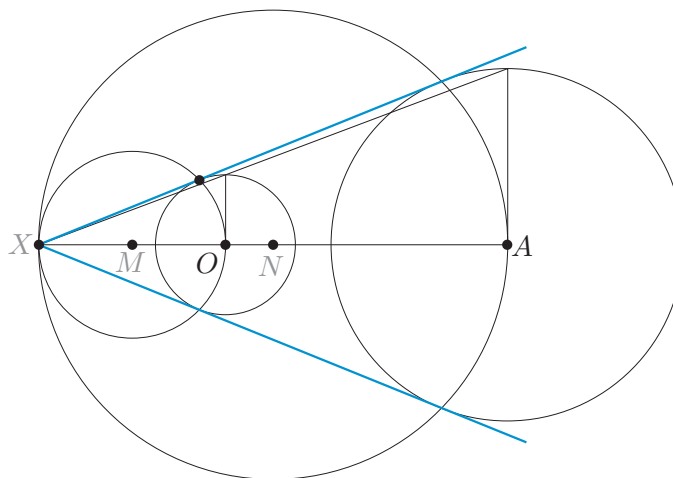
- (c) 1. Draw a line between the centre, O , of the circle and the given point, P . (O1)
2. Construct the midpoint, M , of this line. (C5)
3. Draw the circle with centre at this midpoint that passes through the centre of the original circle. (O2)
4. Draw the line from the intersection, T , of the two circles to the given point. (O1)

This last line is a tangent, since it is at right angles to the radius at the intersection point.



- (d) 1. Draw the line OA joining the centres of the two circles. (O1)
 2. In each circle draw a line through the centre and perpendicular to OA . (C4)
 3. Mark the intersections, on one side of OA , of these perpendiculars with the two circles as B and C respectively.
 4. Draw the line BC and mark the intersection of this line with OA as X . (O1)
 5. Construct the midpoint M of OX . (C5)
 6. Construct the midpoint N of XA . (C5)
 7. Draw the circle through X centre M . (O2)
 8. Draw the circle through X centre N . (O2)
 9. Mark the intersections of these circles with the original circles.
 10. Draw the two lines through X and the two pairs of intersections. (O1)

These two lines will be the required tangents (note that you only need one of the intersections for each tangent).



Solution to Exercise 2.6

Since the set of numbers that are constructible from S is a subfield of $\mathbb{R}$, it must contain $\mathbb{Q}$ by Theorem 1.11 of Chapter 18. The result then follows.

Solution to Exercise 2.7

- (a) The perpendicular bisector of AB passes through the points of intersection of the circles:

$$(x - 1)^2 + (y - 2)^2 = r^2$$

$$(x - 3)^2 + (y - 4)^2 = r^2,$$

where r is the length AB . Subtracting the second equation from the first gives

$$(x - 1)^2 + (y - 2)^2 - [(x - 3)^2 + (y - 4)^2] = 0,$$

which simplifies to $x + y - 5 = 0$.

- (b) The line through AB has equation $2(y - 2) = 2(x - 1)$, that is $y - x - 1 = 0$. To find the required point of intersection we need to solve this equation simultaneously with $y + x - 5 = 0$. This gives $x = 2, y = 3$, as we would expect.

Solution to Exercise 2.8

The trisection of $2\pi/5$ is the angle $2\pi/15$. It is perhaps easier to work using degrees at this stage, so we are asked to construct an angle of 24° given an angle of 72° . This is possible since we can construct an angle of 60° , which means we can then construct an angle of 12° . We can double the angle by copying and so construct an angle of 24° .

Solution to Exercise 2.9

The angle α can be trisected if, and only if, we can construct, over the field $\mathbb{Q}(\cos(\alpha))$, the number $x = \cos \beta$ where $\alpha = 3\beta$. Now

$$\cos \alpha = \cos 3\beta = 4 \cos^3 \beta - 3 \cos \beta,$$

and so x is a root of the polynomial $4x^3 - 3x - \cos(\alpha)$. If this polynomial is irreducible, then x has a minimal polynomial of degree 3 over $\mathbb{Q}(\cos(\alpha))$, and so cannot be constructed. If it is reducible, it has a linear factor and a quadratic factor. The root of the linear factor is constructible as it is just a linear combination of constructible numbers.

Solution to Exercise 2.10

If it were possible to construct a regular 9-gon with ruler and compasses, then it would be possible to construct an angle of $2\pi/9$ radians. We would then be able to bisect this angle to obtain an angle of $\pi/9$ radians. However, in Theorem 2.10 we showed that this is not possible.

Solution to Exercise 2.11

$$(a) \quad x = e^{\frac{2i\pi}{5}} = \cos\left(\frac{2\pi}{5}\right) + i \sin\left(\frac{2\pi}{5}\right)$$

$$\frac{1}{x} = e^{-\frac{2i\pi}{5}} = \cos\left(\frac{2\pi}{5}\right) - i \sin\left(\frac{2\pi}{5}\right)$$

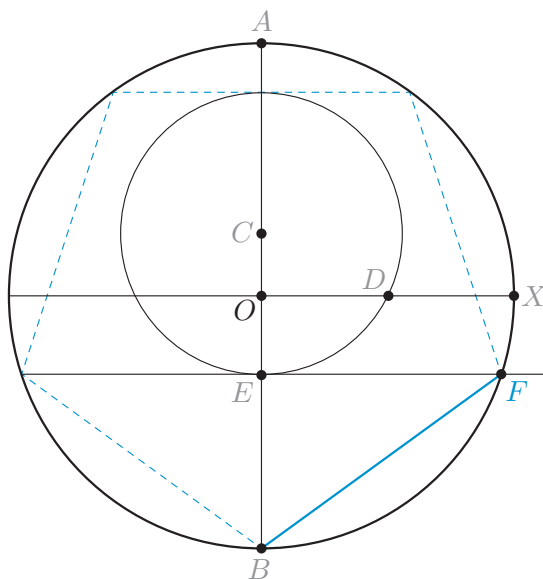
and so $x + \frac{1}{x} = 2 \cos\left(\frac{2\pi}{5}\right) = c$, as required.

Now

$$\begin{aligned} c^2 + c - 1 &= \left(x + \frac{1}{x}\right)^2 + x + \frac{1}{x} - 1 = \frac{1}{x^2}[(x^2 + 1)^2 + x^3 + x - x^2] \\ &= \frac{1}{x^2}[x^4 + 2x^2 + 1 + x^3 + x - x^2] \\ &= \frac{1}{x^2}[x^4 + x^3 + x^2 + x + 1]. \end{aligned}$$

Since $x^4 + x^3 + x^2 + x + 1 = 0$, we have $c^2 + c - 1 = 0$ and so $c = \frac{-1 \pm \sqrt{5}}{2}$. Therefore $\cos\left(\frac{2\pi}{5}\right) = \frac{-1 + \sqrt{5}}{4}$ since the cosine is positive.

1. Draw a circle of centre O and radius 1, and draw a diameter to that circle. Mark the intersections of the diameter with the circle as A and B .
2. Bisect OA twice so as to find the point C with $OC = \frac{1}{4}$.
3. Draw the diameter XY of the circle perpendicular to AB .
4. Find the midpoint D of OX .
5. Draw a circle of centre C and radius CD ($CD = \frac{\sqrt{5}}{4}$).
6. Mark the intersection of this circle with OB as E .
7. Mark the intersection of the perpendicular to OB at E with the original unit circle as F ($OF = \cos\left(\frac{2\pi}{5}\right)$).
8. BF will be one side of the pentagon.



Constructing a regular pentagon

- (b) (i) Suppose that $\cos(\theta) = \frac{1+\sqrt{5}}{4}$.

Then the formula $\cos(2\theta) = 2\cos^2(\theta) - 1$ gives

$$\begin{aligned}\cos\left(\frac{2\pi}{5}\right) &= 2\left(\frac{1+\sqrt{5}}{4}\right)^2 - 1 \\ &= \frac{6+2\sqrt{5}}{8} - 1 \\ &= \frac{-2+2\sqrt{5}}{8} \\ &= \frac{-1+\sqrt{5}}{4}.\end{aligned}$$

The result follows since $\cos\left(\frac{2\pi}{5}\right) = \frac{-1+\sqrt{5}}{4}$.

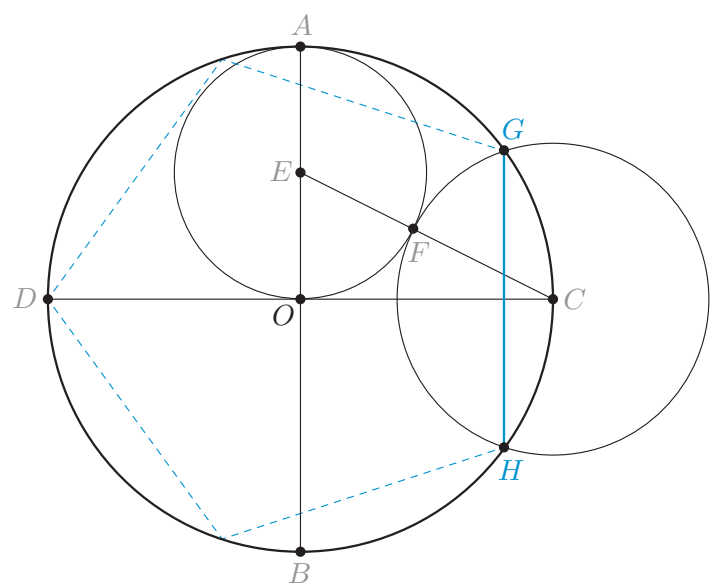
- (ii) The length of a side of a decagon will equal the length of the base of an isosceles triangle with equal sides of length 1 enclosing an angle of $\frac{\pi}{5}$. Thus if the length of the base is c , the cosine rule, $c^2 = a^2 + b^2 - 2ab\cos C$, gives

$$\begin{aligned}c^2 &= 2 - 2\frac{1+\sqrt{5}}{4} \\ &= \frac{6-2\sqrt{5}}{4}.\end{aligned}$$

The result follows since $\left(\frac{\sqrt{5}-1}{2}\right)^2 = \frac{6-2\sqrt{5}}{4}$.

- (iii) 1. Draw a circle of centre O and radius 1, and draw a diameter to that circle. Mark the intersections of the diameter with the circle as A and B .
2. Draw the diameter CD of the circle perpendicular to AB .
3. Bisect OA so as to find the point E with $OE = \frac{1}{2}$.
4. Draw the line CE , which will have length $CE = \frac{\sqrt{5}}{2}$.
5. Draw the circle of centre E and radius EO , and let it meet CE at F , so $CF = \frac{\sqrt{5}-1}{2}$.
6. Draw the circle of centre C and radius CF , and let it meet the original circle at G .

The side CG is one side of a regular decagon, and the construction of the regular pentagon follows.



A second construction of a regular pentagon